



Offre n°2025-08991

Doctorant F/H De quoi re?vent les GNN ? Me?thode d'interpre?tabilit? base?e sur l'extraction de motifs

Type de contrat : CDD

Niveau de diplôme exigé : Bac + 5 ou équivalent

Fonction : Doctorant

A propos du centre ou de la direction fonctionnelle

Le centre Inria de l'Université de Rennes est un des neuf centres d'Inria et compte plus d'une trentaine d'équipes de recherche. Le centre Inria est un acteur majeur et reconnu dans le domaine des sciences numériques. Il est au cœur d'un riche écosystème de R&D et d'innovation : PME fortement innovantes, grands groupes industriels, pôles de compétitivité, acteurs de la recherche et de l'enseignement supérieur, laboratoires d'excellence, institut de recherche technologique.

Contexte et atouts du poste

Dans le cadre d'un partenariat (vous pouvez choisir entre)

- public avec ANR

Mission confiée

Contexte

Les GNN (Graph Neural Network) [1] sont des re?seaux de neurones particulie?rement populaires actuellement car ils permettent de traiter des donne?es de type graphes (ex: re?seaux sociaux, mole?cules, graphes de connaissances,etc.).

Ces approches sont devenues l'e?tat de l'art pour des ta?ches comme la pre?diction de lien ou la classification de graphes, avec des re?sultats impressionnants par rapport aux approches ante?rieures. Toutefois comme les autres approches s'appuyant sur des re?seaux de neurones, les de?cisions des GNN ne sont pas inte?pre?tables : il est quasiment impossible pour un expert humain de comprendre ce qui a conduit a? cette de?cision.

Contrairement aux approches de réseaux de neurones classiques (i.e., prenant en entrée du texte ou des images) pour lesquelles de nombreuses méthodes d'explicabilité existent [2,3], peu de travaux se sont intéressés aux cas des graphes. Dans [4] une approche d'explicabilité pour les GNN utilisant la fouille de motifs [5] a été proposée. Cette méthode a la particularité de s'appuyer directement sur les composantes actives pendant la prise de décision du GNN, extrayant des motifs appelés "règles d'activation". Ces règles d'activation sont ensuite reliées aux données graphes d'entrée, permettant de générer des explications sous forme de sous-graphes.

Cette méthode préliminaire a plusieurs limitations. En premier lieu, les règles d'activation ne sont extraites que pour une couche donnée du GNN, limitant leur expressivité. De plus, les non-activations ne sont pas prises en compte alors qu'elles peuvent être cruciales pour expliquer une décision. L'autre limitation importante est la trop grande quantité de règles d'activation générées, liée à la nature combinatoire de l'approche d'extraction. Il faut donc une mesure de qualité pour sélectionner un sous-ensemble pertinent. La mesure utilisée actuellement par la méthode est basée sur des propriétés statistiques de l'ensemble de règles d'activation, mais ne fait pas le lien entre les règles et les parties des graphes d'entrée auxquelles elles correspondent.

Objectif de la thèse

L'objectif de la thèse est de fournir aux utilisateurs humains des explications riches, précises et compréhensibles des décisions d'un GNN. Dans un premier temps le travail de la thèse se focalisera sur l'expressivité des règles d'activation extraites du GNN. On souhaite ainsi développer une méthode extrayant des règles d'activation de composantes à partir d'un nombre arbitraire de couches du GNN, et prenant en compte aussi bien les activations que les non activations (e.g., patterns négatifs [6]).

Cela conduira à l'exploration d'un espace immense de règles potentielles, l'approche proposée devra donc retourner un petit sous-ensemble des règles les plus pertinentes pour expliquer la décision du GNN. Pour cela, des méthodes basées sur la Théorie de l'Information (en particulier le principe de longueur de description minimale [7]) seront étudiées.

Une dernière contribution théorique sera l'étude des approches de « traduction » de ces règles d'activation expressives vers l'espace des graphes d'entrée, afin de fournir des explications compréhensibles basées sur des éléments des graphes d'entrée. Les applications envisagées sont d'une part les graphes de molécules chimiques, et d'autre part les graphes de connaissance du web sémantique. La piste envisagée pour effectuer une traduction robuste entre les règles d'activation et les graphes d'entrée est d'exploiter la connaissance stockée dans des LLM pour capturer une partie de la sémantique du domaine des graphes d'entrée.

[1] F. Scarselli, M. Gori, A. C. Tsoi, M. Hagenbuchner, G. Monfardini. The Graph Neural Network Model. In IEEE Transactions on Neural Networks, vol. 20, no. 1, pp. 61-80 (2009).

[2] M. Túlio Ribeiro, S. Singh, C. Guestrin. "Why Should I Trust You?": Explaining the Predictions of Any Classifier. KDD 2016: 1135-1144

[3] Scott M. Lundberg, Su-In Lee: A Unified Approach to Interpreting Model Predictions. NIPS 2017: 4765-4774

[4] L. Veyrin-Forrer, A. Kamal, S. Duffner, M. Planetevit, C. Robardet. On GNN explainability with activation rules. Data Min Knowl Disc (2022).

[5] C. Aggarwal, J. Han. Frequent Pattern Mining. Springer, Cham (2014).

[6] T. Guyet, R. Quiniou. NegPSpan: efficient extraction of negative sequential patterns with embedding constraints. Data Min. Knowl. Discov. 34(2): 563-609 (2020)

[7] P. Grünwald. The Minimum Description Length Principle. The MIT Press (2007)

Principales activités

Principales activités :

- Développer des programmes
- Concevoir des plateformes expérimentales
- Rédiger des articles
- Tester, modifier jusqu'à valider
- Diffuser les travaux via des publications et des exposés
- Présenter l'avancée des travaux aux partenaires

Compétences

Le candidat ou la candidate devra avoir une appétence pour le machine learning en général et en particulier les réseaux de neurones, les statistiques, l'algorithmique et la programmation.

Avantages

- Restauration subventionnée
- Transports publics remboursés partiellement
- Congés: 7 semaines de congés annuels + 10 jours de RTT (base temps plein) + possibilité d'autorisations d'absence exceptionnelle (ex : enfants malades, déménagement)
- Possibilité de télétravail (après 6 mois d'ancienneté) et aménagement du temps de travail
- Équipements professionnels à disposition (visioconférence, prêts de matériels informatiques, etc.)
- Prestations sociales, culturelles et sportives (Association de gestion des œuvres sociales d'Inria)
- Accès à la formation professionnelle

Rémunération

Salaire brut : 2200€

Informations générales

- **Thème/Domaine** : Représentation et traitement des données et des connaissances
Systèmes d'information (BAP E)
- **Ville** : Rennes
- **Centre Inria** : [Centre Inria de l'Université de Rennes](#)
- **Date de prise de fonction souhaitée** : 2025-10-01
- **Durée de contrat** : 3 ans
- **Date limite pour postuler** : 2025-08-31

Contacts

- **Équipe Inria** : [LACODAM](#)
- **Directeur de thèse** :
Cellier Peggy / Peggy.Cellier@irisa.fr

A propos d'Inria

Inria est l'institut national de recherche dédié aux sciences et technologies du numérique. Il emploie 2600 personnes. Ses 215 équipes-projets agiles, en général communes avec des partenaires académiques, impliquent plus de 3900 scientifiques pour relever les défis du numérique, souvent à l'interface d'autres disciplines. L'institut fait appel à de nombreux talents dans plus d'une quarantaine de métiers différents. 900 personnels d'appui à la recherche et à l'innovation contribuent à faire émerger et grandir des projets scientifiques ou entrepreneuriaux qui impactent le monde. Inria travaille avec de nombreuses entreprises et a accompagné la création de plus de 200 start-up. L'institut s'efforce ainsi de répondre aux enjeux de la transformation numérique de la science, de la société et de l'économie.

L'essentiel pour réussir

Attention: Les candidatures doivent être déposées en ligne sur le site Inria. Le traitement des candidatures adressées par d'autres canaux n'est pas garanti.

Consignes pour postuler

Déposer en ligne CV et lettre de motivation

Sécurité défense :

Ce poste est susceptible d'être affecté dans une zone à régime restrictif (ZRR), telle que définie dans le décret n°2011-1425 relatif à la protection du potentiel scientifique et technique de la nation (PPST). L'autorisation d'accès à une zone est délivrée par le chef d'établissement, après avis ministériel favorable, tel que défini dans l'arrêté du 03 juillet 2012, relatif à la PPST. Un avis ministériel défavorable pour un poste affecté dans une ZRR aurait pour conséquence l'annulation du recrutement.

Politique de recrutement :

Dans le cadre de sa politique diversité, tous les postes Inria sont accessibles aux personnes en situation de handicap.